Check for updates

ARTIFICIAL INTELLIGENCE

"Data will solve robotics and automation: True or false?": A debate

Nancy M. Amato[1], Seth Hutchinson[2], Animesh Garg[3], Aude Billard[4], Daniela Rus[5], Russ Tedrake[5], Frank Park[6], Ken Goldberg[7]*

Copyright © 2025 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works

Leading researchers debate the long-term influence of model-free methods that use large sets of demonstration data to train numerical generative models to control robots.

INTRODUCTION

Nancy M. Amato, Seth Hutchinson, and Ken Goldberg

The robotics and automation research community is experiencing what historian of science Thomas Kuhn would call a "paradigm shift" (*1*). This was the subject of a featured debate at the IEEE International Conference on Robotics and Automation (ICRA), the largest research conference for the field, where many presentations celebrated algorithmic ("model-based") breakthroughs over the decades. However, since 2012, advances in deep "neural" networks that combine unprecedented amounts of example data, advances in stochastic gradient descent techniques, and advances in computing, in particular GPUs (graphics processing units), have produced remarkable results in computer vision and speech recognition. And with the emergence of ChatGPT and associated large vision language models (VLMs) in 2022, with associated advances in natural language processing and denoising diffusion methods, the paradigm of "end-to-end" (also known as "model-free") approaches to perception and control suggest an entirely new, data-driven approach to robotics and automation.

Some researchers believe, at one extreme, that the data-driven paradigm is overrated, and some at the other extreme believe that model-based approaches are outdated and should be retired. Many researchers take less extreme positions, believing both paradigms have merit and that the model-based and model-free paradigms can coexist (*2*).

As cochairs of the 2025 IEEE International Conference on Robotics and Automation (ICRA), held in Atlanta GA, USA, 26 to 29 May 2025, N.M.A. and S.H. decided to host a conference-wide keynote debate in the last session of the conference. To moderate the debate, we invited K.G., who has been very active in the past 40 years of model-based robotics and automation research and also active in robot learning (recently elected president of the Robot Learning Foundation). K.G. formulated a catchy title and short description: "Will the future of robotics and automation be written in code or in data? Is the *Handbook of Robotics* obsolete? Are home humanoids overhyped? Join us for a high-voltage debate about physics vs pixels, theory vs terabytes."

The three of us then worked to identify a panel of six debaters, inviting three to represent each side in the spirit of the classic Oxford debates. Six highly respected researchers accepted our invitation: A.G., D.R., and R.T. advocating for "true," and A.B., Leslie Kaebling (of MIT), and F.P. advocating for "false." We asked each debater to take extreme positions and prepare 3-min introductory presentations, responses to questions, and closing remarks.

The debate was very well attended—more than 1000 audience members cheering or jeering as debaters alternated between the two positions. The debaters focused on software and control and did not address the many unsolved issues related to hardware, materials, and design, nor did they have time to discuss how such massive datasets can be collected and processed.

Interestingly, most of the debaters conveyed a shared respect for hybrid methods that ultimately combine model-free and model-based methods. Ultimately, a show of hands indicated that not many audience members were persuaded to change their initial positions, but everyone was pushed to think more deeply about the new data paradigm.

DATA ARE NOT MERELY BENEFICIAL; RATHER, THEY ARE INDISPENSABLE AND FOUNDATIONAL

Animesh Garg

Humanity has frequently developed "know-how" before the "know-why." We engineered technologies mainly using observational understanding—such as metalworking, sailing ships, internal combustion engines, and airplanes (all foundational to civilization)—yet preceding a thorough understanding of their underlying sciences.

We are in the midst of building next-generation technology—"computational and physical intelligence" with data-driven empirical methods. Modern computing has only recently emerged as a tool for scientific discovery, yet it fundamentally redefines both the acquisition of know-how and the pursuit of know-why. We can now create and collect planetary-scale data and perform comprehensive computations to compress them into a single model. This unprecedented speed and scale has upended our approach to scientific inquiry compared with any other point in human history.

In our quest to understand intelligence and "solve" robotics, I contend that data are not merely beneficial; rather, they are indispensable and foundational. This necessity stems from several critical perspectives:

[1]University of Illinois, Urbana, IL, USA. [2]Northeastern University, Boston, MA, USA. [3]Georgia Institute of Technology, Atlanta, GA, USA. [4]École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland. [5]Massachusetts Institute of Technology, Cambridge, MA, USA. [6]Seoul National University, Seoul, South Korea. [7]University of California, Berkeley, Berkeley, CA, USA.
*Corresponding author. Email: Goldberg@berkeley.edu

Downloaded from https://www.science.org on September 05, 2025

The debaters. Top row (left to right): Nancy M. Amato, Seth Hutchinson, and Ken Goldberg. Bottom row (left to right): Animesh Garg, Aude Billard, Russ Tedrake, and Frank Park.

Downloaded from https://www.science.org on September 05, 2025

1) General-purpose intelligence, particularly in physical robots, is ambiguous, open ended, and underspecified. Data offer guidance for achieving distributional similarity where a closed-form solution for common sense is absent. Moreover, self-supervised learning—through observation or experience—allows for the encoding of concepts without manual programming. Last, learning with sparse yet critical feedback enhances robustness in observed behaviors and facilitates the discovery of novel ones.

2) Accurate models foster innovation. Conversely, incomplete models can hinder the discovery of true optima, because these might be unrepresentable or unattainable with existing optimization tools. Our understanding of generalizable autonomy in robotics is currently insufficient to define even necessary conditions, let alone sufficient ones.

3) Robotics represents the evolution of computing, moving beyond desktop, mobile, and cloud environments into the physical world. Whereas data-driven approaches are often criticized for lacking elegance, the empirical process of data collection, creation, and curation is inherently a scientific endeavor. Moreover, the construction of complex technologies like foundation models involves intricate processes, despite their seemingly monolithic outputs. Therefore, we must apply principles of abstraction from other computing domains to build sophisticated yet well-behaved robotic systems.

4) Data-driven methods have successfully unified diverse problems in fields like language processing and computer vision. However, robotics remains a collection of disparate communities. A potential solution lies in adopting the foundation model perspective, which could unify many seemingly unrelated problems within the robotics community, given that all robots are governed by the same fundamental laws of physics.

Generalizable autonomy for physical systems is an open-ended problem that requires us to find any necessary solution before we can develop a more elegant, interpretable, and efficient approach. Our endeavors in creating data-driven technological artifacts are merely one step in enhancing our comprehension of the essence of intelligence itself.

MODELS GIVE MEANING TO DATA AND ARE ESSENTIAL FOR UNDERSTANDING ROBOT CONTROL

Aude Billard

All science is grounded in data, but data alone are insufficient. I do not believe that data alone will "solve" robotics and automation.

Consider astronomy: It started off with the simple observation of the sky. But thousands of people look at the sky each day, and millions have before. Although they all marvel at its beauty, most of these people, if asked, would be incapable of explaining the reason why the stars are aligned as they are or distinguishing stars from planets. Astronomers from ancient civilizations ago did not use these millions of human eye observations to decipher the underlying laws of physics. Rather, they used few but meticulously gathered data points. Most importantly, they constantly went back and forth from building hypothetical models to data to test their theories.

Some argue that what was true for hardcore science does not apply to robotics. Robotics, and its deployment through large-scale automation, is an engineering discipline, relying on experimentation and numerical models tailored to the specific platforms being controlled.

Robotics today seems like an impatient child who wants it all, now and at all costs. It dreams of deploying a magic tool that will control a variety of robotic bodies from flying to walking and swimming machines and endow them with large language models' level of cognition in the flesh, enabling them to interact seamlessly with all humans, from children to the elderly, and to operate outdoors and indoors in all weather, with minimal changes to our homes and habits.

If this is what is meant by solving robotics, then the task may seem unsolvable with a traditional scientific approach, given that none of the sciences we know today can capture, through mathematical laws, all of the world's complexity—let alone the complexity of human beings. It is hence understandable to seek an alternative road.

The success of large language models relies on the fact that humans had put at the disposal of computers billions of pieces of information—from entire encyclopedias to largely uncurated material—all of which provided a view of a large chunk of human knowledge that can be put into words. Should robotics approach the problem in a similar way to achieve similar levels of knowledge, it needs to encapsulate all human and other animals' knowledge of perceiving and doing things. This would require embodying those creatures. However, existing interfaces for transferring knowledge from humans to robots are still cumbersome and intrusive and offer only a crude approximation of human sensing, and we are still a long way from embodying birds, fishes, or worms.

Yet, suppose robotics does succeed in this endeavor, but only through sheer computational power applied to vast datasets, without the development of any underlying general theoretical principles. Then, the [Springer] *Handbook of Robotics* would not become obsolete. On the contrary, it might become our only cherished resource, because we may no longer be able to produce a third edition, or any new handbook for that matter, beyond coding guidelines for how best to use the latest artificial intelligence (AI) toolbox.

We started by contrasting data from models, debating which may matter more. Perhaps what matters is neither alone but rather the interplay between gathering data and refining models that is essential to building an understanding of how robots can act in the world. Failing this, we may just be building another tower of Babel.

I BELIEVE THAT WE NEED BOTH DATA AND MATHEMATICAL MODELS

Daniela Rus

Although I was assigned to argue in favor of data "solving" robotics and automation, my actual position is more nuanced: I believe that we need both data and mathematical models. Robotics has long drawn strength from first principles. Physics gives us elegant models for simple tasks: grasping a rigid object, stabilizing a robot arm, or following a collision-free path. When the environment is structured, the materials known, and the goals well defined, analytical models deliver precision and insight. The moment we leave the lab and enter the real world, those assumptions collapse. Grasping becomes "pick the right object out of a cluttered drawer." Navigation becomes "move through a crowded room without bumping into people—or social norms." In these domains, first principles alone are no longer enough.

The challenge is complexity. Real-world tasks are multimodal, context-dependent, and filled with ambiguity. They involve soft, deformable objects; occluded or partial views; unpredictable humans; and environments that shift over time. There is no single equation for how to unload a dishwasher in a new home or to hand a tool to someone who may—or may not—reach out to take it. To function in these settings, robots need more than structure—they need experience. And experience comes from data.

Data allow us to capture the richness and complexity of the real world—complexity that first-principles models alone cannot fully express. They provide breadth, variation, and direct engagement with reality. Through data, we can model tasks that are too intricate to formalize, capture interactions too subtle to simulate, and learn behaviors too diverse to program by hand. Data enable robots to build representations that reflect the ambiguity and variability of real environments—accounting for uncertainty, human intent, and the fluidity of physical execution. Most critically, data allow robots to respond intelligently when things go wrong—not by following a prescripted rule, but by drawing on prior experience. They do not recover because we told them what to do; they recover because they have seen something like it before.

Data also bring resilience and enable robots to operate under noisy perception, ambiguous instructions, and changing dynamics. With enough structured experience, AI systems can learn to filter uncertainty, infer intent, and adapt online. Just as large language models gained generality from vast, varied text, robots will require diverse, richly labeled physical experience to achieve robust intelligence—not just more video, but multisensory, embodied data: joint torques, contact forces, hand trajectories, and human gazes.

That is why, in our lab, we have built a physical kitchen task testbed that goes far beyond passive observation. We instrument people with sensors that record body pose, muscle activity, force interactions, and gaze patterns as they perform real tasks—cutting, pouring, reaching, and handing off. These recordings generate dense, multimodal datasets that capture the what and the how of human movement. From this, we train AI systems that learn to perform those same tasks—gracefully, robustly, and with the flexibility to adapt when the environment changes.

As I noted above, I also believe that mathematical models and first principles still matter. They help structure the learning process, provide constraints, and enforce physical consistency. But without data, we cannot capture the long tail of variation that defines the real world. The path to capable, general-purpose robots is not paved solely in equations—it is built on experience. To solve robotics and automation, we need systems that combine the rigor of physics with the richness of data. That is how we move from brittle to robust, from reactive to adaptive, from lab demos to real-world autonomy.

Ultimately, solving robotics and automation requires a marriage of disciplines. The rigor of physics gives us constraints and structure. The breadth of data gives us coverage and flexibility. One without the other is incomplete. If we want robots that thrive not just in factories but also in homes, hospitals, and everyday human environments, we must invest in the data that capture the richness of those worlds. The future of robotics will be driven by equations alongside experience.

MANIPULATION REQUIRES "COMMON SENSE"; LARGE DATA AND LARGE MODELS CAN PROVIDE IT

Russ Tedrake

I do believe that the next most impactful steps toward "solving" robotics lie in large-scale data collection and large pretrained models. How could I not? The advances in large models for language, vision, and now video are undeniable. For the last few years at Toyota Research Institute (TRI) and other labs, I have seen more and more examples of robots programmed via imitation learning performing dexterous manipulation tasks that would have been impossible to imagine even a few years ago; most recently, we have empirically validated that multitask pretraining improves robustness and/or allows us to add new skills with fewer demonstrations and that performance predictably increases as the scale and diversity of the pretraining data grow (*3*, *4*).

Why are data so fundamental to this success? Large-scale pretraining from diverse multitask data is the best (only?) way that I know to program "common sense." ChatGPT gets this from reading all of the text on the internet, and our robots are slowly getting physical common sense from seeing enough demonstrations of manipulation. Most people agree that large models are useful for manipulation tasks that require common-sense language understanding (e.g., "pick up the extinct animal"), but many people underestimate how essential common sense is for low-level control.

For example, at TRI, we recently taught a bimanual robot with simple grippers to core and slice an apple using kitchen tools. No two apples are the same, and the diversity of rollouts that we observed in this experiment

Downloaded from https://www.science.org on September 05, 2025

was impressive. Most impressive to me were the subtle recovery maneuvers that the robot could make when apple pieces would slip or squirt around on the cutting board. Having spent enough time myself manually programming state machines for tasks like this, it is incredibly powerful to see how quickly we can add recovery behaviors to the demonstration dataset to improve the performance; as these recovery maneuvers start to appear automatically through pretraining via transfer from demonstrations in completely different tasks, we start to have what I would call common sense for dexterous manipulation. This suggests a clear path to true open-world robustness.

For me, "solving" robotics is a very long-term agenda. Well after we have general-purpose robots out in the open world performing useful work, we will still have a lot more work to do to deeply understand the laws of robotics. We are in the early phases, where everything feels messy. I am optimistic that we will quickly deepen our empirical understanding of these new capabilities and that we will eventually have theorems that guide us. However, precisely because the world is so messy and because so much of manipulation runs through common sense, I think we will need large-scale data and algorithms that can digest them, in order for humans (or machines) to build the right simple models to help us deeply understand.

A MORE GROUNDED STRATEGY IS NEEDED

Frank Park

There is no doubt that multimodal foundation models represent a transformational moment in human history, but expecting a parallel revolution in robotics is premature at best and wishful thinking at worst. Robotics is a different beast entirely: Real-world data are scarce, simulations remain unreliable, and robots in their many embodiments must contend with an endless variety of environments and tasks. Real-world performance requirements can be unforgiving, to say nothing of the massive computing, energy, and data demands of today's foundation models.

None of this is to say that foundation models for robotics are not worth pursuing—they are, but a more grounded strategy is needed. Increasingly, machine learning researchers are questioning whether scaling deep learning alone is enough; Goyal and Bengio (*5*) have argued that better inductive biases (priors, preferences, or assumptions) are the key to human-level cognition and its corresponding hierarchy of representations. I believe the same is true for robotics.

Fortunately, in robotics, we already have a wealth of such inductive biases. Biology offers insights into how humans control movement through layers of abstraction, as well as the interplay between feedforward versus feedback and internal versus external coordinates. Physics provides models of motion, force, and compliance. Geometry helps to simplify complex movements into lower-dimensional representations and to exploit any underlying symmetries. Programming gives us tools like motion description languages that formalize manipulation tasks in a device-independent manner. Current robot foundation models do not sufficiently exploit these inductive biases.

Instead of collecting endless hours of video footage on how to clean a bathroom or assemble a chair, we would do better to study tasks where the nuances of control are everything—like how a chef slices a carrot or a cucumber. What matters is not just the motion, but the tactile feedback at the fingertips, the compliance in the wrist and elbow, and the complex forces involved at each stage. Manipulation tasks like these are precisely the kinds of challenges where traditional model-based methods fall short; foundation models trained with the right data can truly make a difference.

Let us also not forget: Convolutional neural networks (CNNs) and transformers were not some serendipitous discovery of blind scaling. They were designed by researchers who understood the importance of symmetry and equivariance (translation for CNNs, permutations for transformers). Robotics deserves the same level of intentional design. Rather than force-fitting models developed for vision and language to robots, new models that capture the inductive biases inherent to robotics are needed. It is time we treated robotics not as an extension of deep learning but as its own grand challenge.

REFERENCES AND NOTES

1. T. Kuhn, *The Structure of Scientific Revolutions* (Univ. Chicago Press, 1962).
2. K. Goldberg, How to close the 100,000 year data gap in robotics and automation? *Sci. Robot.* **10**, eaea7390 (2025).
3. R. Tedrake, "Multitask transfer in TRI's large behavior models for dexterous manipulation," Stanford seminar, 25 April 2025; available at https://youtube.com/watch?v=TN1M6vg4CsQ.
4. TRI LBM Team, J. Barreiros, A. Beaulieu, A. Bhat, R. Cory, E. Cousineau, H. Dai, C.-H. Fang, K. Hashimoto, M. Zubair Irshad, M. Itkina, N. Kuppuswamy, K.-H. Lee, K. Liu, D. McConachie, I. McMahon, H. Nishimura, C. Phillips-Grafflin, C. Richter, P. Shah, K. Srinivasan, B. Wulfe, C. Xu, M. Zhang, A. Alspach, M. Angeles, K. Arora, V. Campagnolo Guizilini, A. Castro, D. Chen, T.-S. Chu, S. Creasey, S. Curtis, R. Denitto, E. Dixon, E. Dusel, M. Ferreira, A. Goncalves, G. Gould, D. Guoy, S. Gupta, X. Han, K. Hatch, B. Hathaway, A. Henry, H. Hochsztein, P. Horgan, S. Iwase, D. Jackson, S. Karamcheti, S. Keh, J. Masterjohn, J. Mercat, P. Miller, P. Mitiguy, T. Nguyen, J. Nimmer, Y. Noguchi, R. Ong, A. Onol, O. Pfannenstiehl, R. Poyner, L. Priebe Mendes Rocha, G. Richardson, C. Rodriguez, D. Seale, M. Sherman, M. Smith-Jones, D. Tago, P. Tokmakov, M. Tran, B. Van Hoorick, I. Vasiljevic, S. Zakharov, M. Zolotas, R. Ambrus, K. Fetzer-Borelli, B. Burchfiel, H. Kress-Gazit, S. Feng, S. Ford, R. Tedrake, A careful examination of large behavior models for multitask dexterous manipulation. arXiv:2507.05331 [cs.RO] (2025).
5. A. Goyal, Y. Bengio, Inductive biases for deep learning of higher-level cognition. *Proc. R. Soc. A* **478**, 20210068 (2022).

Acknowledgments: This debate ("Data Will Solve Robotics and Automation: True or False?") was held at the 2025 IEEE International Conference on Robotics and Automation. A video of the debate can be found at https://youtu.be/PfvctjoMPk8?si=Zwpt-QfJ3EiqferR.

10.1126/scirobotics.aea7897

Downloaded from https://www.science.org on September 05, 2025